

UNITED STATES PATENT APPLICATION

FOR

**A METHOD AND APPARATUS FOR SHARING MEMORY SPACE ACROSS
MULTIPLE PROCESSING UNITS**

INVENTOR:

Ravikrishna Cherukuri

Prepared by:

Blakely, Sokoloff, Taylor & Zafman
12400 Wilshire Boulevard
Seventh Floor
Los Angeles, California 90025
(408) 720-8598

Attorney's Docket No. 004906.P026

"Express Mail" mailing label number: EL 485756981US

Date of Deposit: December 18, 2000

I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to the Assistant Commissioner for Patents, Washington, D. C. 20231

Clara Wallin

(Type or printed name of person mailing paper or fee)

(Signature of person mailing paper or fee)

December 18, 2000

(Date signed)

A METHOD AND APPARATUS FOR SHARING MEMORY SPACE ACROSS MUTLIPLE PROCESSING UNITS

FIELD OF THE INVENTION

5 The invention relates to the field of computer processing and memory accessing. More specifically, the invention relates to accessing of multiple memory units by multiple processing units.

BACKGROUND OF THE INVENTION

10 With the continued increase in high-speed processing by various processing units, the need for higher-speed accessing of memory units located "off-chip" of such processing units has become increasingly critical. One example where such computer processing has become increasingly critical is in the field of telecommunications. Network elements are located across and within different networks in the
15 telecommunications industry to process or switch all of the different packets that are received. In particular, a given network element includes a number of different ports to receive and transmit packets of data therefrom. Accordingly, upon receipt of a given packet of data on a port, a network element determines which port to transmit this packet out from. In a typical network element, there are a number of different line cards having
20 a number of different ports that are each receiving and/or transmitting packets of data. These network elements must, therefore, be able to buffer and process or switch these packets at very high data rates.

 Accordingly, certain network elements incorporate multiple processors and multiple memory units in order to process or switch all of these data packets being
25 received. Different network elements incorporate different configurations for the incorporation of multiple processors and multiple memory units. In one such configuration, a network element establishes a one-to-one relationship between a given

memory unit, such as a dynamic random access memory (DRAM), and a given processor such that temporary storage of data for this processor is provided by this particular memory unit. However, this one-to-one configuration between a processor and a memory unit may not allow for the optimal accessing of data from the memory units, as a
5 given processor may only access one memory unit during operation.

In an alternative configuration, a network element allows for sharing of these multiple memory units by the multiple processors. Accordingly, one processor can access data from any of the multiple memory units. However, this configuration can also be problematic as bottlenecking may occur during the accessing of the different memory
10 units when multiple processors may try to access data from the same memory unit at the same time.

SUMMARY OF THE INVENTION

A method and apparatus for sharing memory space of multiple memory units by multiple processing units are described. In an embodiment, a method includes storing a set of data across more than one of at least two memory units upon determining that the number of sets of data is static. The method also includes storing the set of data within a single memory unit of the at least two memory units upon determining that the set of data is dynamic.

In another embodiment, a method for accessing a set of data from a number of memory units includes reading the set of data from a single memory unit from the number of memory units, upon determining that the set of data is dynamic. The method also includes reading the set of data from any of the number of memory units, upon determining that the set of data is static.

In one embodiment, an apparatus includes a first memory unit to store a set of static data and a first set of dynamic data. The apparatus also includes a second memory unit coupled to the first memory unit. The second memory unit stores the set of static data and a second set of dynamic data. Additionally, the apparatus includes a first processing unit coupled to the first and second memory units. The first processing unit can read the set of static data from the first memory unit or the second memory unit. Moreover, the apparatus includes a second processing unit coupled to the first and second memory units and the first processing unit. The second processing unit can read the set of static data from the first memory unit or the second memory unit.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention may be best understood by referring to the following description and accompanying drawings which illustrate such embodiments.

In the drawings:

5 **Figure 1** is block diagram illustrating a system that incorporates embodiments of the present invention;

Figure 2 illustrates portions of in-ring network elements 102-108, according to embodiments of the present invention;

Figure 3 illustrates a block diagram of ingress packet processing circuitry 212a-d or egress packet processing circuitry 214a-d, according to embodiments of the present invention;

Figure 4 is a flowchart for storage of data within memory units 310-314, according to embodiments of the present invention;

Figure 5 illustrates a block diagram of memory units 310-314 having stored therein both dynamic and static sets of data, according to embodiments of the present invention; and

Figure 6 is a flowchart for accessing or reading of data within memory units 312-318, according to embodiments of the present invention.

DETAILED DESCRIPTION

A method and apparatus for sharing memory space of multiple memory units by multiple processing units are described. Embodiments of the present invention are described herein in terms of a multiple processing units and multiple memory units inside of a network element for the processing or switching of data packets within a network. However, embodiments of the present invention are not so limited, as such embodiments may be incorporated into other types of systems for other types of processing.

Additionally, portions of the detailed description describe “packets” in terms of Internet Protocol (IP) packets. However, embodiments of the present invention are not so limited, as other types of packet data can be incorporated into embodiments of the present invention. In particular, the term “packet” may include, but is not limited to,
5 Asynchronous Transfer Mode (ATM) cell, Frame Relay packets and voice over IP packets. Moreover, in the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be evident, however, to one skilled in the art that the present invention may be practiced without these specific details.

10 **Figure 1** is block diagram illustrating a system that incorporates embodiments of the present invention. In particular, Figure 1 illustrates system 100 that includes network ring 114, which is comprised of in-ring network element 102, in-ring network element 104, in-ring network element 106 and in-ring network element 108. System 100 also include non-ring network element 110, non-ring element 111 and non-ring network
15 element 112, which are coupled to network ring 114 through in-ring network element 102, in-ring network element 104 and in-ring network element 106, respectively. In an embodiment, non-ring elements 110-112 are routers. However, embodiments are not so limited, as non-ring elements 110-112 can be any other type of network element that can transmit and/or receive Internet Protocol (IP) packets. For example, non-ring elements
20 110-112 could be a switch or bridge within a telecommunications network.

In one embodiment, the connection among in-ring network element 102, in-ring network element 104, in-ring network element 106 and in-ring network element 108 allow for bi-directional traffic. Accordingly, this bi-directional capability allows for redundancy in the communication between the different network elements, such that if a
25 given line of communication is lost, the data traffic to be transmitted thereon can be rerouted in the opposite direction to reach its intended destination within the ring architecture.

In an embodiment, system 100 transmits data traffic among the different network elements, both in-ring and non-ring, employing the Synchronous Optical Network (SONET) standard or Synchronous Digital Hierarchy (SDH). In one embodiment, data traffic among in-ring network element 102, in-ring network element 104, in-ring network element 106 and in-ring network element 108 includes both Time Division Multiplexing (TDM) traffic as well as Internet Protocol (IP) traffic within a same SONET signal. A further description of the operation of system 100 and the network elements therein is described in more detail below.

The architecture and configuration of system 100 is by way of example and not by way of limitation, as embodiments of the present invention can be incorporated in other types of systems. For example, other such systems could incorporate less or more network elements into the network ring and/or network elements attached thereto. Moreover, embodiments of the present invention are not limited to the network ring architecture as illustrated in Figure 1. Examples of other types of network architectures that can incorporate embodiments of the present invention include, but are not limited to, a point-to-point configuration, point-to-multipoint configuration and/or a hub configuration.

Figure 2 illustrates portions of in-ring network elements 102-108 (for purposes of Figure 2, hereinafter “network element 102”), according to embodiments of the present invention. As shown, network element 102 includes line cards 202a-d and control card(s) 220, such that control card(s) 220 are coupled to each of line cards 202a-d. The number of line cards illustrated are for the sake of simplicity and not by way of limitation, as a lesser or greater number of line cards can be included within network element 102. Additionally, network element 102 includes packet mesh 226, which includes a full mesh such that each of line cards 202a-d are coupled to one another. For example, line card 202a is coupled to line cards 202b-d through packet mesh 226. However, embodiments of the present invention are not limited to a full mesh for the transmission of packets

among line cards 202a-d, as any type of switching method that switches based on the addressing scheme described herein can be incorporated into embodiments of the present invention. For example, in one embodiment, line cards 202a-d could be coupled together using a switch fabric, such that the line cards are coupled to a control card, which
5 provides for the switching therein.

Line cards 202a-d include physical connection circuitry 210a-d, ingress packet processing circuitry 212a-d, egress packet processing 214a-d and TDM processing circuitry 216a-d, respectively. Physical connection circuitry 210a-d can be coupled to lines external to network element 102, as shown, which can carry optical and/or electrical
10 signals. In one embodiment, line cards 202a-d of network element 102 may be connected to an optical line transmitting SONET OC-N signals. Moreover, in an embodiment, line cards 202a-d of network element 102 may be connected to an electrical line such as a T1, T3, E1, E3, Ethernet, Gigabit Ethernet, etc.

In an embodiment, each line card 202a-d can be coupled to four optical and/or
15 electrical lines. In another embodiment, each line card 202a-d can be coupled to eight optical and/or electrical lines. However, embodiments of the present invention are not so limited, as a lesser or greater number of optical and/or electrical lines can be coupled to network element 102 through line cards 202a-d. Additionally, physical connection circuitry 210a-d are coupled to ingress packet processing circuitry 212a-d, respectively,
20 such that IP packet data being received from the optical and/or electrical lines is passed from physical connection circuitry 210a-d to ingress packet processing circuitry 212a-d, respectively.

Ingress packet processing circuitry 212a-d is coupled to packet mesh 226. Accordingly, each ingress packet processing circuitry 212a-d is coupled to each egress
25 packet processing circuitry 214a-d, respectively, on other line cards 202a-d through packet mesh 226. Moreover, egress packet processing circuitry 214a-d is respectively coupled to physical connection circuitry 210a-d, such that IP packet data traffic coming

in from packet mesh 226 from ingress packet processing circuitry 212a-d is transmitted from egress packet processing circuitry 214a-d to physical connection circuitry 210a-d, respectively.

With regard to the TDM data traffic, a switch fabric is formed among line cards 202a-d and control cards 220, as illustrated by the dashed lines in Figure 2. In particular, physical connection circuitry 210a-d is coupled to TDM processing circuitry 216a-d, respectively, for the receiving and transmitting of TDM data traffic into and out of network element 102. Moreover, each of TDM processing circuitry 216a-d is coupled to control cards 220, such that TDM data traffic transmitted among line cards 202a-d is switched through control cards 220.

Figure 3 illustrates a block diagram of ingress packet processing circuitry 212a-d or egress packet processing circuitry 214a-d, according to embodiments of the present invention. In particular, Figure 3 includes array of processing units 302 along with memory unit 312, memory unit 314, memory unit 316 and memory unit 318. Additionally, array of processing units 302 includes processing unit 304, processing unit 306, processing unit 308 and processing unit 310.

The number of processing units and memory units within the block diagram of Figure 3 is by way of example and not by way of limitation. In particular, as shown by the "... " within the block diagram of Figure 3, ingress packet processing circuitry 212a-d and egress packet processing circuitry 214a-d can include any number of processing units and memory units therein, according to embodiments of the present invention.

Moreover, each of processing units 304-310 is coupled to each of memory units 312-318, thereby allowing for communication between any given processing unit and any given memory unit. In particular, processing unit 304 is coupled to memory units 312-318, while processing unit 306 is coupled to memory units 312-318. Similarly, processing unit 308 is coupled to memory units 312-318, while processing unit 310 is also coupled to memory unit 312-318.

In an embodiment wherein array of processing units 302 and memory units 312-318 are incorporated into a network element, each of processing units 304-310 includes a packet descriptor cache. In particular, as illustrated by Figure 3 processing unit 304 includes packet descriptor cache 320, and processing unit 306 includes packet descriptor
5 cache 322. Additionally, processing unit 308 includes packet descriptor cache 324, and processing unit 310 includes packet descriptor cache 326.

The operation of array of processing units 302 and memory units 312-318 will now be described in more detail in conjunction with Figures 4 and 5. **Figure 4** is a flowchart for storage of data within memory units 312-318, according to embodiments of
10 the present invention. Additionally, **Figure 5** illustrates a block diagram of memory units 312-318 having stored therein both dynamic and static sets of data, according to embodiments of the present invention. As shown in Figure 5, memory units 312-318 include different sets of data stored therein. These sets of data can include one to any number of bits.

In particular, memory unit 312 includes first set of static data 502, first set of
15 dynamic data 504 and second set of static data 512. Additionally, memory unit 314 includes first set of static data 502, second set of dynamic data 506 and second set of static data 512. Memory unit 316 includes first set of static data 502, third set of dynamic data 508 and second set of static data 512. Further, memory unit 318 includes first set of
20 static data 502, fourth set of dynamic data 510 and second set of static data 512. These sets of data within memory units 312-318 will now be described in more detail in conjunction with Figure 4.

In particular, Figure 4 illustrates method 400 that commences when a unit, such as an Application Specific Integrated Circuit (ASIC) or other processing unit located
25 external to ingress packet processing circuitry 212a-d or egress packet processing circuitry 214a-d, determines to store a set of data within one of memory units 312-318. For example, returning to Figure 2, physical connection circuitry 210a-d could store or

write data packets received from outside network elements 102-108 into any of memory units 312-318. In further illustration, one of ingress packet processing circuitry 212 could forward the data packet to one of egress packet processing circuitry 214 through packet mesh 226, thereby causing data packets to be written or stored into memory units 312-318 of egress packet processing circuitry 214. Moreover, in another embodiment, such a unit, which stores sets of data into memory units 312-318, can be one of processing units 304-310.

Upon determining to store a set of data within memory units 312-318, a unit determines whether the set of data is dynamic, at process decision block 402. In an embodiment, a “dynamic” set of data can be associated with specific types of data that are stored in memory units 312-318. For example when array of processing elements 302 and memory units 312-318 are within a network element for data packet transmission as described above in conjunction with Figures 1 and 2, data packets can be defined as dynamic while other types of data, such as an IP address forwarding table, stored in memory units 312-318 can be considered static (i.e., non-dynamic).

In one embodiment, multiple units, either internal or external to ingress packet processing circuitry 212a-d or egress packet processing circuitry 214a-d, can store sets of data within memory units 312-318. In one such embodiment, a first unit may only store a given set of data, such as dynamic data, in memory units 312-318, while a second unit may only store a different set of data, such as static data, in multiple of memory units 312-318. Accordingly, in such an embodiment, the decision of whether a set of data is stored within one or multiple memory units 312-318 is dictated by the unit that is performing the store. For example, returning to Figures 1 and 2, physical connection circuitry 210a-d, a first unit, may only store data packets as dynamic sets of data, into memory units 312-318, while control card(s) 220, a second unit, may only store table and configuration information, such as an IP address table, as static sets of data into memory units 312-318.

In another embodiment, a “dynamic” set of data is defined to include data that does not remain in memory units 312-318 for more than a predetermined threshold. This predetermined threshold is configurable based on the type of data that is being incorporated into embodiments of the present invention.

5 For example, in an embodiment wherein array of processing elements 302 and memory units 312-318 are within a network element for data packet transmission as described above in conjunction with Figures 1 and 2, a set of data that remains within memory units 312-318 for less than a predetermined threshold is a data packet that is received for transmission through the network to which the network element is attached.

10 Such a data packet is typically buffered within memory units 312-318 until one of processing units 304-310 processes the packet to determine the packet’s transmission, such that the packet is forwarded based on the address associated with this data packet. For example, one of processing units 304-310 determines the packet’s destination based on an Internet Protocol (IP) address-forwarding table. A data packet is by way of example of a dynamic set of data and not by way of limitation, as any other type of data
15 that does not remain within memory units 312-318 for a predetermined threshold can also be defined as dynamic.

In an embodiment, if the set of data is dynamic, the unit stores this data into one of memory units 312-318, at process block 404. Returning to Figures 1 and 2, in an
20 embodiment, a given data packet being transmitted and/or received within system 100 is broken into a number of sets of data, which are stored in memory units 312-318. In one such embodiment, a given set of data includes 256 bytes. Moreover, in an embodiment, the unit generates a packet descriptor that is placed into each of packet descriptor caches 320-326 within processing units 304-310. This packet descriptor includes a pointer that
25 points to the head of a given data packet. Moreover, when a given data packet is across more than a single set of data within memory units 312-318, a preceding set of data includes a pointer to the subsequent set of data, thereby allowing processing units 304-

310 to locate the data packet within memory units 312-318. In one embodiment, the different sets of data that comprise a given data packet can be stored across memory units 312-318.

Returning to Figure 5 to help illustrate, assume that a given data packet stored in memory units 312-318 includes, in the following order, third set of dynamic data 508, second set of dynamic data 506, and fourth set of dynamic data 510. Accordingly, a packet descriptor stored in packet descriptor caches 320-326 of processing units 304-310 includes a pointer for this given data packet that points to third set of dynamic data 508. Additionally, third set of dynamic data 508 includes a pointer that points to the location of second set of dynamic data 506. Similarly, second set of dynamic data 506 includes a pointer that points to the location of fourth set of dynamic data 510. Accordingly, when processing this given data packet, one of processing units 304-310 accesses or reads third set of dynamic data 508 based on the packet descriptor stored in packet descriptor caches 320-326 within processing units 304-310 and follows the pointers to the subsequent sets of data associated with this data packet (i.e., second set of dynamic data 506, and fourth set of dynamic data 510). In an embodiment wherein the data packet is an IP packet, the beginning of the data packet includes a header that enables one of processing units 304-310 to forward or switch the data packet based on an IP address contained therein.

The above-described example illustrated the storing of a set of dynamic data within only one of memory units 312-318. However, embodiments of the present invention are not so limited. In another embodiment, such data could be stored in two of memory units 312-318. For example, returning to Figure 5, first set of dynamic data 504 could be stored in both memory unit 312 and memory unit 314.

In one embodiment, if the set of data is not dynamic, such data is considered static and is stored into each of memory units 312-318, at process block 406. To further illustrate, Figure 5 shows first set of static data 502 stored in memory unit 312, memory unit 314, memory unit 316 and memory unit 318. Additionally, second set of static data

512 is stored across memory unit 312, memory unit 314, memory unit 316 and memory unit 318. However, embodiments of the present invention are not limited to the storage of static data in all of memory units 312-318. For example, in another embodiment, sets of static data can be stored in one-half of the memory units. In an alternative
5 embodiment, sets of static data can be stored in any two of the memory units.

Additionally, any of processing units 304-310 can also read sets of data that were written or stored in any of memory units 312-318. **Figure 6** is a flowchart for accessing or reading of data within memory units 312-318, according to embodiments of the present invention. Figure 6 is described with reference to the reading or accessing of sets
10 of data by processing unit 304. Embodiments of the present invention are not limited to the reading or accessing of sets of data by processing units 304-310. For example, units, such as physical connection circuitry 210a-d, external to ingress packet processing circuitry 212a-d or egress packet processing circuitry 214a-d can read or access the sets of data from memory units 312-318, according to embodiments of the present invention.

Method 600 of Figure 6 commences when data is going to be read or accessed by
15 processing unit 304. Processing unit 304 determines the location of the set of data to be read from within memory units 312-318, at process block 602. In an embodiment wherein a data packet is being accessed, processing unit 304 determines the beginning location of this data packet using a packet descriptor that is placed into packet descriptor cache 320 of processing unit 304, as described above. In particular, the packet descriptor
20 for this data packet includes the location of the data packet within memory units 312-318. As described above, a given data packet can be stored across multiple sets of data that can be stored across the different memory units 312-318.

This determination of a location for accessing a set of data from memory units
25 312-318 is by way of example and not by way of limitation, as other techniques can be employed for the determining of this location. For example, an instruction to be executed by processing unit 304 can identify this location, as is known in the art. Moreover, in an

embodiment, the sets of data being accessed by processing unit 304 can be either static or dynamic. Accordingly, the static set of data can be stored across multiple memory units 312-318, thereby having multiple locations within such memory units. Processing unit 304, therefore, may have determined multiple locations from which to access the set of data from.

Processing unit 304 determines whether the set of data to be read or accessed is dynamic, at process decision block 604. Upon determining that the set of data is dynamic, processing unit 304 reads the set of data from the given memory unit from among memory units 312-318 wherein the data is stored, at process block 606.

Conversely, upon determining that the set of data is non-dynamic (i.e., static), processing unit 304 attempts to access the set of data from memory unit 312. Processing unit 304 determines whether memory unit 312 is available to access the set of data therefrom, at process decision block 608. Memory unit 312 could be unable access the set of data from memory unit 312 for different reasons. One reason may be that another device, such as processing unit 306, may be accessing data from memory unit 310 at the time that processing unit 304 is trying to access the set of data. Another reason may be that memory unit 312 may not be operating such that no devices can access sets of data therefrom.

Upon determining that memory unit 312 can be accessed, processing unit 304 accesses the set of data from memory unit 312, at process block 610. However, if that memory unit 312 is unavailable for accessing of the set of data, processing unit 304 determines whether memory unit 314 is available to access the set of data therefrom, at process decision block 612. Upon determining that memory unit 314 can be accessed, processing unit 304 accesses the set of data from memory unit 314, at process block 614. However, if that memory unit 314 is unavailable for accessing of the set of data, processing unit 304 determines whether memory unit 316 is available to access the set of data therefrom, at process decision block 616. Upon determining that memory unit 316

can be accessed, processing unit 304 accesses the set of data from memory unit 316, at process block 618. However, if memory unit 316 is unavailable for accessing of the set of data, processing unit 304 determines accesses the set of data from memory unit 318, at process block 620.

5 Figure 6 illustrates the accessing of a set of static data beginning with memory unit 312, then memory unit 314, followed by memory unit 316 and then memory unit 318. This order of traversal is by way of example and not by way of limitation, as embodiments of the present invention can employ any other order of traversal of memory units that have stored a set of the static data.

10 As illustrated, embodiments of the present invention provide for the storage of sets of data across multiple memory units, such that not every set of data is stored in every memory unit, while providing storage across multiple memory units for those sets of data that are more likely to remain within such memory units for a longer period in comparison to other sets of data. Accordingly, embodiments of the present invention preclude bottlenecking at a given memory unit, as sets of data are copied across multiple
15 memory units. Additionally, embodiments of the present invention provide a configuration such that not every set of data is required to be copied into every memory unit.

 The operation for the storage of the data within memory units 312-318 is
20 described in terms of processing units 304-310 performing this task. However, embodiments of the present invention are not so limited, as other units are capable of performing this storage of data within memory units 312-318. For example, a processing unit that is external to ingress packet processing circuitry 212a-d or egress packet processing circuitry 214a-d could store data into memory units 312-318, according to
25 embodiments of the present invention.

 Memories described herein include a machine-readable medium on which is stored a set of instructions (i.e., software) embodying any one, or all, of the

methodologies described herein. Software can reside, completely or at least partially, within these memories and/or within processors and/or storage devices. For the purposes of this specification, the term "machine-readable medium" shall be taken to include any mechanism that provides (i.e., stores and/or transmits) information in a form readable by
5 a machine (e.g., a computer). For example, a machine-readable medium includes read only memory (ROM); random access memory (RAM); magnetic disk storage media; optical storage media; flash memory devices; electrical, optical, acoustical or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); etc.

Thus, a method and apparatus for sharing memory space of multiple memory
10 units by multiple processing units have been described. Although the present invention has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the invention. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive
15 sense.